



# Limpieza y transformación de datos con Python

**Servicio Nacional de Aprendizaje - SENA**  
Nivel de formación: **Complementaria**

## 01 Presentación

Estimado aprendiz, bienvenido a este espacio de formación denominado “Limpieza y transformación de datos con Python”, el cual tiene como fin desarrollar las competencias, habilidades y los conocimientos necesarios para poder seleccionar y preparar los datos necesarios para realizar un adecuado análisis en Python, teniendo en cuenta los requerimientos técnicos y las necesidades organizacionales.

¿Sabían que la limpieza de los datos es quizás la tarea más importante para garantizar el análisis efectivo de los datos? Así es, esta tarea ocupa aproximadamente el 80 % del tiempo requerido para la implementación de proyectos de analítica de datos, por ende, este programa de formación será de mucha utilidad para aquellas personas interesadas en emprender en el mundo de los datos.

Inscríbase en: [www.senasofiaplus.edu.co](http://www.senasofiaplus.edu.co)



**Código**  
21710113



**Horas**  
48



**Duración**  
1 mes



**Modalidad**  
Virtual



## 02 Justificación del programa

Los algoritmos tradicionales generalmente presentan elementos estáticos, lo que significa que a partir de los mismos datos de entrada, siempre ofrecen los mismos resultados de salida. Sin embargo, la analítica de datos, como parte de la inteligencia artificial, introduce nuevas formas de trabajo que buscan patrones y características emergentes en los datos a medida que cambian con el tiempo, adaptándose a diversas situaciones. Aunque esta capacidad pueda parecer mágica, es importante destacar que estos procesos, en los que los algoritmos aprenden, solo tienen sentido gracias a la calidad de los datos de entrada.

A pesar de los avances en algoritmos, estos carecen de la intuición que los seres humanos poseen. El éxito de los algoritmos depende exclusivamente de los datos que se les proporcionan. Por lo tanto, la selección cuidadosa, limpieza y transformación de los datos son de vital importancia para que los algoritmos puedan descubrir patrones de manera efectiva.

Para dar respuesta al reto anterior, el SENA propone el curso complementario en limpieza de datos que hace parte del módulo enfocado en la ciencia de datos (*Big data*) y tiene como fin dar a conocer los métodos y las herramientas para la limpieza y transformación de los datos, de tal manera que las personas que accedan al programa puedan hacer aportes de valor dentro del sector productivo, aplicando elementos que permitan el aprovechamiento de datos para la toma de decisiones estratégicas y el desarrollo e implementación de diferentes procesos de manera más organizada y con resultados mucho más efectivos.

## 03 Competencia a desarrollar

**220501114** - Sistematizar datos masivos de acuerdo con métodos de analítica y herramientas tecnológicas

## 04 Perfil de Ingreso

### Conocimientos en:

- Manejo en la manipulación de grandes volúmenes de datos *Big data*.
- Manejo de hojas de cálculo.
- Manejo en lenguaje de programación en Python (variables, estructuras básicas).
- Nivel básico de inglés.
- Recopilación de información de acuerdo con los requerimientos técnicos.
- Identificación de objetivos del objeto de análisis.
- Identificación de fuentes de datos.

### Académicos:

- Grado 9 aprobado.
- Requisito adicional: superar prueba de aptitud, motivación, interés y competencias mínimas de ingreso.
- Requisito de ley: edad mínima definida en la ley (14 años).

## 04 Estrategia Metodológica

Centrada en la construcción de autonomía para garantizar la calidad de la formación en el marco de la formación por competencias, el aprendizaje por proyectos y el uso de técnicas didácticas activas que estimulan el pensamiento para la resolución de problemas simulados y reales; soportadas en la utilización de las Tecnologías de la Información y la Comunicación - TIC, integradas en ambientes virtuales de aprendizaje, que en todo caso recrean el contexto productivo y vinculan al aprendiz con la realidad cotidiana y el desarrollo de las competencias.

Igualmente, debe estimular de manera permanente la autocrítica y la reflexión del aprendiz sobre el quehacer y los resultados de aprendizaje, esto se logra a través de la vinculación activa de las cuatro fuentes de información para la construcción de conocimiento:

- El instructor - tutor.
- El entorno.
- Las TIC.
- El trabajo colaborativo.